RESEARCH ARTICLE

Analysis of agricultural products' market demand and price fluctuation trend using big data

Weili Liu*

School of Electronic Information Engineering, City University of Zhengzhou, Zhengzhou, Henan, China.

Received: April 18, 2024; **accepted:** August 19, 2024.

In a globalized economic environment, demand and price fluctuations in agricultural markets have significant economic and social implications. The study used big data analytics and machine learning algorithms to predict agricultural market demand and price fluctuations. By collecting and collating multi-source data including trading, climate, and consumer behavior, a predictive model was built to analyze the key factors affecting agricultural prices. The model evaluation showed that climatic conditions and market demand were the main drivers of price fluctuations. Sensitivity analysis further revealed the relative importance of these factors. The results of the study had practical implications for producers, policy makers, and market analysts to formulate strategies. However, the generalization ability and data quality of the model were the main limitations of the study, and future studies should be expanded on this basis to obtain wider applicability and higher accuracy.

Keywords: big data analysis; agricultural markets; price fluctuation; demand dynamics.

***Corresponding author:** Weili Liu, School of Electronic Information Engineering, City University of Zhengzhou, Zhengzhou 452370, Henan, China. Email: liu_weili@outlook.com.

Introduction

In a globalized economic environment, demand and price fluctuations in agricultural markets have significant economic and social implications. The study of product markets has always been an important part of the field of economics and market analysis. This market is highly dynamic and influenced by multiple factors such as climate change, policy adjustments, technological advances, and changing consumer preferences. The supply and price fluctuations of agricultural products have profound implications for the global economy, food security, and the well-being of farmers. Therefore, accurate analysis and prediction of market demand and price fluctuations of agricultural products are essential for formulating effective agricultural policies, optimizing supply chain management, and ensuring food safety.

In the current research context, big data and advanced analytical methods play an important role in the optimization of agricultural marketing channels. Zhai *et al*. explored a path to optimize agricultural marketing channels based on innovative industry chains, which highlighted the potential of supply chain management in enhancing market efficiency [1]. Xiong discussed the marketing strategy of agricultural products under the background of "Internet plus" and stressed the importance of digital platforms in promoting marketing innovation [2]. Meanwhile, deep learning is also becoming increasingly important in analyzing Internet marketing of agricultural products as demonstrated by Liu *et al*. in their research [3]. Regarding the research on market demand and consumer behavior, Nie *et al*. analyzed the market demand and government regulation of China's agricultural product quality classification system, providing an important perspective for understanding consumer preferences and policy impact [4]. In addition, big data analysis plays a key role in revealing the coupling relationship between agricultural product marketing and agricultural economic development in the context of "Internet plus" [5]. In terms of marketing strategies, the previous studies emphasized the application of artificial intelligence and deep learning technology in formulating marketing strategies for agricultural products, highlighting the value of technology in improving the accuracy of marketing strategies [6, 7]. Further, Huang *et al*. confirmed the impact of the use of big data analytics in smart agriculture on agricultural commerce and product marketing, providing data-driven insights for agricultural practices and policy making [8], which showed that big data and analytical tools played a central role in understanding and predicting market dynamics for agricultural products, providing new strategies and solutions for effective marketing of agricultural products. Although many studies have explored the applications of big data in agricultural markets, there are still challenges in data quality, model selection, and policy development in market analysis, which are important areas for future research to continue to explore. Most of the existing studies focus on a single data source or a specific market, which lacks comprehensive analysis of multi-source data. In addition, how to improve the accuracy and generalization ability of prediction models is also an important issue in current research.

The main purpose of this study was to use big data technology to deeply analyze and predict the demand and price fluctuations in the agricultural market. Specifically, the research aimed to develop and validate a set of predictive models that accurately reflected market dynamics, and, through these models, provided insight into the key factors influencing market volatility including, but not limited to, environmental factors, policy changes, consumer behavior, and global market trends. Big data analysis and machine learning techniques were adopted in this study. This research not only focused on developing efficient forecasting tools, but also aimed to deepen the theoretical basis for understanding market dynamics and explore new ways to apply big data technology in the field of economics. The results of this study could provide practical insights to producers, policymakers, and market analysts to help them better manage the risks posed by market volatility. By continuing to explore and optimize these approaches, professionals could respond more effectively to the constant changes and challenges in the global agricultural market.

Materials and methods

Data sources and collection methods

The data collection strategy aimed to obtain a broad and relevant data set to ensure a comprehensive understanding of agricultural market demand and price fluctuations. Data collection of this study included transaction data, climate data, consumer behavior data, and more from multiple data sources obtained through a variety of collection methods to ensure the comprehensiveness and accuracy of the data. Market transaction data was one of the key data sources, mainly from agricultural exchanges, government agriculture departments, and market research reports. The data included volume, transaction prices, and market supply and demand and was collected by regularly downloading data from exchange and government websites, as well as subscribing to market research reports. The climate and environmental data were important for the analysis of agricultural markets, which was obtained from the meteorological office, environmental monitoring stations, and satellite data, covering temperature, rainfall and extreme

Table 1. Data supplier information.

Supplier Name	City	Province	Country
GreenFarm Co.	Beijing	Beijing	China
FreshHarvest	Shanghai	Shanghai	China
AgriSupplies Ltd	Guangzhou	Guangdong	China
EcoProduce Inc.	Chengdu	Sichuan	China
NaturalFoods	Wuhan	Hubei	China

events such as droughts or floods. The consumer behavior data was obtained from retailers, e-commerce platforms, and social media analysis, which included purchase history, consumer preferences, and social media trends, through partnerships with retailers and e-commerce platforms, while using social media analytics tools for data collection. In addition, policy and economic indicators data including policy changes, economic growth rate, and inflation rate were collected from government announcements and economic research institutions by monitoring government announcements and subscribing to economic research reports. The technological progress data came from agricultural technology research reports and professional journals, covering new agricultural technologies and improved crop yields. By synthesizing the above data sources and collection methods, this study was able to construct a comprehensive and detailed dataset to provide a solid basis for analyzing market demand and price fluctuations of agricultural products. These data not only help to understand the dynamics of the market, but also provide accurate input for model construction and forecasting. The data source information was listed in Table 1.

Data sorting and quality control

The collected data was cleaned, transformed, and standardized to ensure data quality and suitability. Data collation and quality control were the key steps to ensure the accuracy of analysis. After data collection, all data was reviewed to identify missing, duplicate, or inconsistent data records using statistical description analysis and outlier detection algorithm. The data was further cleaned to resolve errors and inconsistencies using deleting or interpolating missing values, correcting faulty data records, identifying and handling outliers before transformed into a format or structure suitable for analysis using data encoding, variable conversion, and normalization. The data now was ready to be integrated from different sources to build a consistent dataset using data matching and merging and resolving inconsistencies between data sources. Eventually, the data was standardized to make sure the data follows the same standard, especially time series data, which included date format unification, currency unit conversion, and so on. A total of 3,000 pieces of data that covered the period from 2018 to 2023 and the major agricultural producing areas in China including Beijing, Shanghai, Guangdong, Sichuan, and Hubei provinces was included in this study. These regions, due to their geographic and climatic diversity, could provide a rich sample of data to better reflect the demand and price fluctuations of agricultural markets across the country. The various indicators including temperature, rainfall, purchases, GDP growth, and inflation over the five-year period were covered in collected data. 70% of the data (2,100 records) was used for model training, while the remaining 30% of the data (900 records) was used for model evaluation and performance improvement.

Construction method of prediction model

The machine learning algorithms including linear regression, time series analysis, random forests, and neural networks were applied for model construction. The appropriate model was selected based on the data characteristics and prediction targets. ARIMA model was chosen for time series data, while random forests or neural networks were used for nonlinear relationships and high-dimensional data. The features from the data that had an impact on the predicted target were then selected before model training process using selected features and algorithms, which involved setting model parameters, cross-validation, and so on. The model was evaluated for its predictive power using test datasets. Commonly used evaluation indexes included mean square error (MSE) and determination coefficient (R^2). A linear regression model to predict the price was constructed as shown in equation (1).

$$Price = \beta_0 + \beta_1 \times Temp + \beta_2 \times Rainfall + \beta_3 \times Demand + \epsilon \quad (1)$$

where β_0 was the intercept. β_1 to β_3 were the coefficients related to temperature, rainfall, and market demand. ϵ was the error term. The coefficient estimation and model were then obtained as follows.

$$Price = 50 + 5 \times Temp - 2 \times Rainfall + 0.01 \times Demand \quad (2)$$

In practical applications, the study performed these calculations with more complex model diagnosis and optimization which included normality of the error terms, homoscedasticity, and the fit of the model to check the assumptions of the model. In addition, the robustness of the model was verified with additional datasets to ensure its performance on unknown data.

Model evaluation and performance improvement

The model evaluation applied cross-validation and other evaluation metrics to assess the accuracy and reliability of the model, which involved using statistical measures to evaluate the accuracy, consistency, and reliability of the model, and further tuning the model to improve performance. The performance evaluation involved the common evaluation indicators including MSE, R^2, accuracy, and recall. Among them, MSE and R^2 were commonly used evaluation metrics for regression models. MSE measured the average error between the predicted and actual values of the model, while R^2 measured the proportion of variability explained by the model. The cross verification was performed by dividing the dataset into multiple parts, taking turns using one part for testing and the rest for training, which helped to avoid overfitting and evaluate the model's performance on unknown data. The model parameters were adjusted thereafter, or different algorithms were tried based on the evaluation results, which included changing the complexity of the model, optimizing feature selection, or using different preprocessing methods. The error was calculated using MSE as shown in equation (3).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(ActualPrice_i - PredictedPrice_i)^2 \quad (3)$$

In addition, an R^2 value was calculated to assess the explanatory power of the model as below.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(ActualPrice_i - PredictedPrice_i)^2}{\sum_{i=1}^{n}(ActualPrice_i - \overline{ActualPrice})^2} \quad (4)$$

where $\overline{ActualPrice}$ was the average of the actual price. The model was then adjusted based on the evaluation results. Through this continuous evaluation and adjustment, the predictive performance of the model could be gradually improved. The validation of the model was critical to the ability to generalize unseen data. To ensure the reliability and accuracy of the model, several common model validation methods and strategies were adopted including hierarchical cross-validation, a method of dividing the data into several subsets with each subset taking turns as a test set and the rest as training set; the introduction of validation set that introduced an additional validation set in addition to the training set and the test set for intermediate evaluation of the model to tune the model parameters before final testing and avoid overwriting; and tuning hyperparathyroidism, a strategy to find the best model configuration by changing the hyperparathyroidism of the model, which could improve model performance. By combining these methods and strategies, the generalization ability of the model could be effectively evaluated, and the necessary optimization and adjustment of the model could be made, which not only improved the performance of the model in practical applications, but also ensured its stability and reliability on different datasets. In practice, these validation methods and strategies were usually selected based on the characteristics of the dataset, the complexity of the model, and the available computing resources.

Several techniques were employed to optimize model performance. Feature engineering involved optimizing feature selection and excluding unimportant features, which included feature transformation and combination. Regularization techniques such as L1 and L2 regularization were applied to avoid overwriting. Additionally, integrated learning methods like random forest and gradient boosting were used to enhance model stability and accuracy. Hierarchical cross-validation was used to evaluate the linear regression model built earlier with performance measured by cross-validated MSE and R^2. If MSE was high or R^2 was low, the optimization strategies such as feature reevaluation should be implemented, where each feature's contribution was analyzed, and less influential features were removed or transformed. Parameterize adjustment involved fine-tuning model parameters including adjusting the regularization coefficient or learning rate. The model complexity adjustment involved increasing or decreasing the model's complexity based on data characteristics and model performance.

Results and discussion

The relationship between temperature, rainfall, and market demand

The results showed that the temperature of studied area was from 24 to 27°C with rainfall ranged from 70 to 90 mm and purchasing ranged from 450 to 600 kg. Meanwhile, GDP growth remained at 3.5%, while inflation was 2.1%. The linear regression model for the relationship between temperature, rainfall, and market demand was shown in Figure 1. The data was used to train and validate the predictive models of this study to estimate how different factors affected the market prices of agricultural products. Each characteristic was used as an input variable to predict the price. The predicted and actual values of the model were shown in Figure 2. For this dataset, MSE was calculated as follows.

$$MSE = \frac{1}{5}[(200-205)^2 + (220-215)^2 + (210-210)^2 + (205-200)^2 + (215-220)^2]$$

Interpretation and application of analysis results

After completing the construction and evaluation of the model, the analysis results of the model were interpreted, which involved turning the model's statistical findings into deep insights into market behavior. A linear regression model as equation (1) above was used to predict agricultural prices and the following model parameters (coefficients) were obtained as $\beta_0 = 50$, $\beta_1 = 5$, $\beta_2 = -2$, $\beta_3 = 0.01$. The key findings of this model and their market significance were shown in Table 2. With these parameters, it was possible to understand how different factors affected the price of agricultural products. For example, a warming climate might lead to higher prices due to the lower yields of certain crops in warmer conditions. A negative coefficient of rainfall might indicate that too much rainfall was bad for crop growth, which in turn affected supply. The demand coefficient showed that the increase in market demand had a direct positive effect on the price. These findings were crucial for market participants. Agricultural producers can adjust planting strategies and pricing decisions based on this information, while policymakers can better understand and predict market changes and

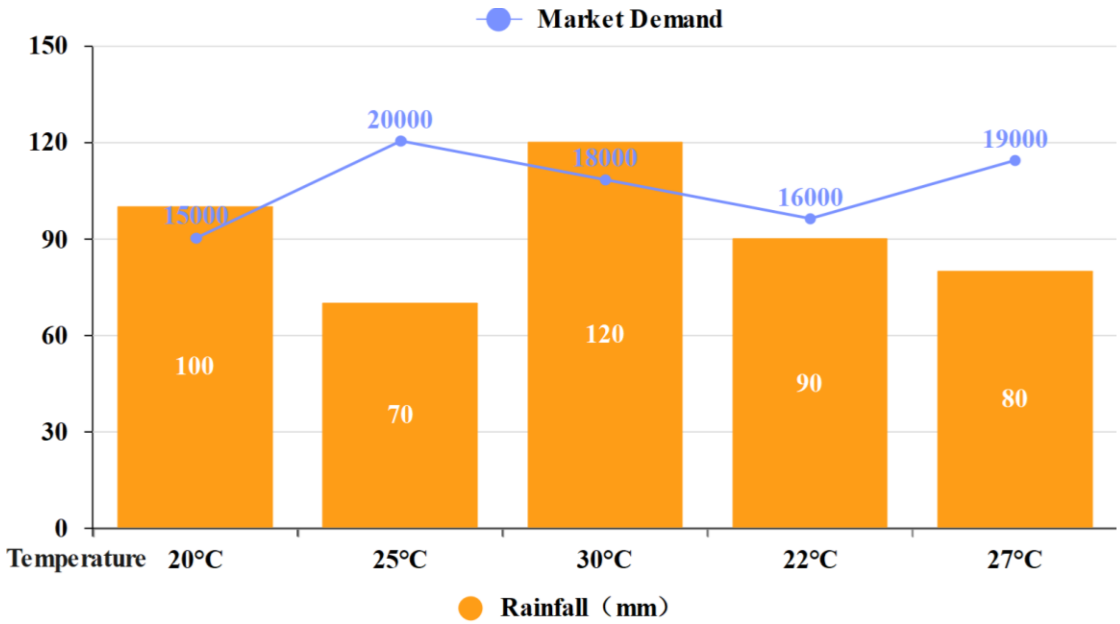

Figure 1. The relationship between temperature, rainfall, and market demand.

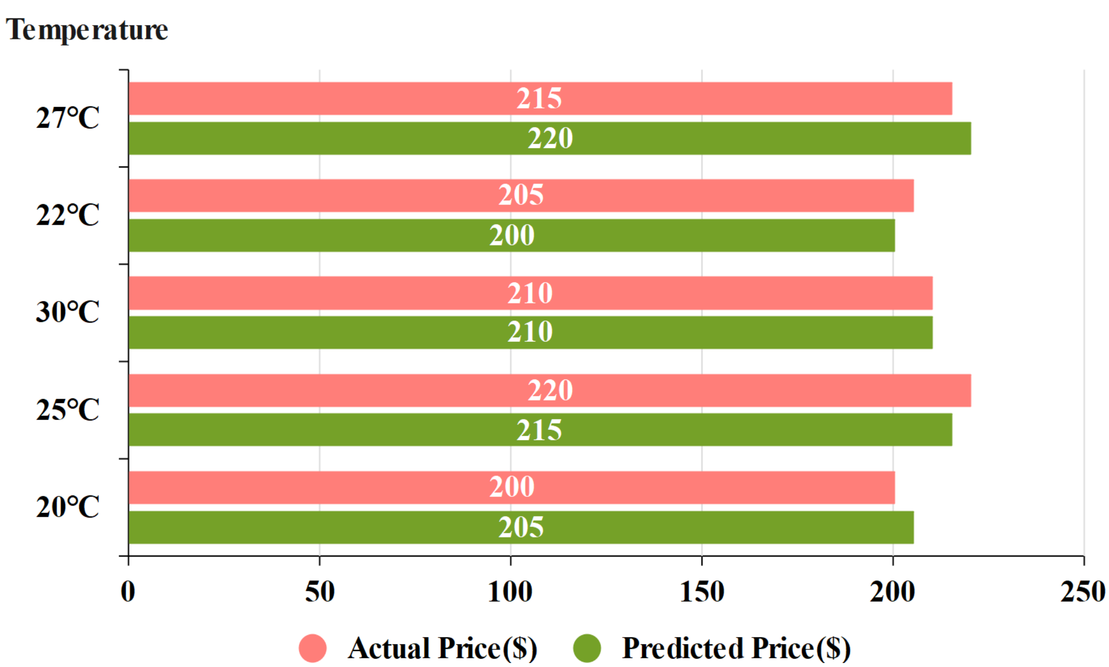

Figure 2. Model evaluation of predicted and actual values.

formulate agricultural and trade policies accordingly. In addition, this analysis can help investors and analysts predict market trends and formulate investment strategies.

Analysis of model sensitivity and key factors

Sensitivity analysis helps to understand the degree to which the model responds to fluctuations in different input variables, thereby revealing the importance of these variables to

Table 2. Market significance of the model.

Argument	Estimated value	Market significance
Intercept β_0	50	Base price, that is, the starting price, other things being equal.
Temperature coefficient β_1	5	For every 1°C increase, the price is expected to increase by 5 units. Reflects the impact of climate on prices.
Rainfall coefficient β_2	-2	For every additional 1mm of rain, the price is expected to decrease by 2 units. May be related to crop growing conditions.
Demand coefficient β_3	0.01	For every additional unit of demand, the price goes up by 0.01 unit. Reflect the impact of market demand.

the predicted results. Sensitivity analysis was conducted based on the predicted data (Figure 3).

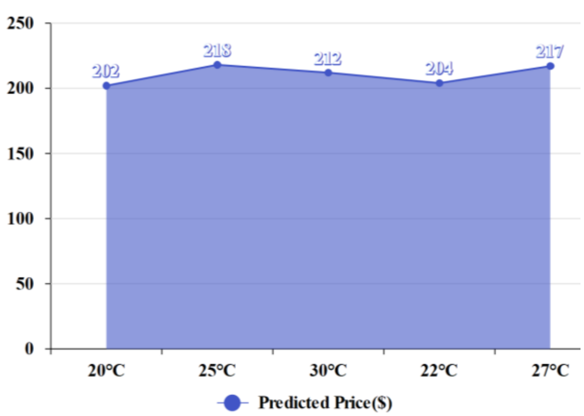

Figure 3. Prediction dataset.

To perform the sensitivity analysis, the study considered the effect of independent changes in each variable on the price as follows.

For temperature (Temp): $$\frac{\partial Price}{\partial Temp} = \beta_1 = 4$$

For rainfall: $$\frac{\partial Price}{\partial Rainfall} = \beta_2 = -3$$

For market demand: $$\frac{\partial Price}{\partial Demand} = \beta_3 = 0.02$$

The sensitivity analysis results were shown in Table 3. The results indicated that temperature and rainfall had a more significant impact on agricultural prices than market demand. This sensitivity analysis was crucial for producers and distributors when planning planting and pricing strategies, especially when climate change and environmental factors were taken into account. In addition, this information was also crucial for policy makers to better understand market dynamics and formulate corresponding agricultural policies and market regulation measures. This sensitivity analysis is very valuable for policy makers and market participants. It can help them understand how agricultural prices are likely to move under different market conditions and environmental variables and develop risk management strategies accordingly. For example, if prices were known to be very sensitive to rainfall, producers and supply chain managers could prepare in advance for possible price fluctuations when predicting the upcoming rainy season.

Existing problems and solutions

When conducting big data analysis of agricultural market demand and price fluctuations, several issues may arise. The data quality and integrity can be compromised by missing values, outliers, or inconsistencies, affecting analysis accuracy. To address this issue, advanced data cleaning and teleprocessing methods such as interpolating missing values and identifying and handling outliers should be applied. Model overwriting

Table 3. Sensitivity analysis results.

Variable	Sensitivity	Explanation
Temperature	4	For every 1°C increase in temperature, the price rises by 4 units, indicating that the price is more sensitive to temperature changes.
Rainfall	-3	For every 1mm increase in rainfall, the price drops by 3 units, indicating that the price is sensitive to the change in rainfall.
Demand	0.02	For every 1,000 units increase in demand, the price rises by 2 units, indicating that the price is relatively insensitive to changes in demand.

Table 4. Typical problems and solutions.

Problem type	Specific problem description	Solution Strategy Example
Data quality	There are a lot of missing values in the data	Use interpolation methods such as k - nearest neighbor (KNN) or multiple interpolation to fill in missing values.
The model is overfitted	High model complexity leads to overfitting	Regularization terms (e.g. L1 regularization) are introduced to punish model complexity.
Multicollinearity	Temperature and rainfall are highly correlated	PCA was used for dimensionality reduction to reduce the correlation between variables.
Generalization ability	Models do not perform well on new data	The stability of the model on different data is improved by using integrated methods such as random forest.

ccurs when the model performs well on training data but poorly on unseen data, which can be mitigated by introducing regularization methods (L1, L2 regularization), performing cross-validation, or simplifying the model structure. Further, multidisciplinary between variables, where explanatory variables are highly correlated, can affect model stability and explanatory power. To reduce linearity, variable selection or conditionality reduction techniques such as principal component analysis (PCA) should be used. The generalization ability of the model may be limited, making it difficult to adapt to new market data or different market environments, which can be improved by using more diverse and representative training data or by trying more adaptive models such as ensemble learning models (Table 4). By identifying and solving these problems, the accuracy of the analysis and the usefulness of the model can be significantly improved. This is of great significance for accurately predicting the demand and price fluctuations of the agricultural products market and providing effective decision support for market participants.

Conclusion

Using big data technology, this study deeply analyzed the demand dynamics and price fluctuations of agricultural products market, revealing the complexity and variability of market behavior. By building and evaluating forecast models, the study successfully identified key factors affecting agricultural prices including climatic conditions, market demand, and supply chain changes. In addition, the sensitivity of the model was analyzed, providing a deeper understanding of market changes. This study provided a new perspective for agricultural market analysis, showing how big data and machine learning techniques could be used to parse complex market dynamics. The findings of this study enriched traditional market analysis methods and provided a comprehensive analysis framework that combined data-driven and theory-driven. However, there were some limitations to the study, which included that the quality and integrity of the data had an important impact on the accuracy of the model, and challenges in data collection and pre-processing remained. The generalization ability of the model needed to be validated and optimized in a broader market context. Future research should explore more diverse data sources as well as more advanced analytical models to further improve the accuracy and reliability of predictions. This study provided valuable insights for market participants, policy makers, and researchers. Agricultural producers and supply chain managers can use these findings to optimize production and pricing strategies and respond to market volatility. Policymakers can use these results to shape more effective agricultural policies and market interventions. In addition, this study provided a solid foundation for future academic research on which to further explore the application of big data in agricultural market analysis. This study has achieved important results in both theory and practice, providing a new perspective and tool for understanding and predicting the demand and price fluctuations of agricultural products market. By continuing to explore and optimize these approaches, we can respond more effectively to the constant changes and challenges in the global agricultural market.

Acknowledgements

This study was supported by Key Scientific Research Project Plan of Colleges and Universities in Henan Province (Grant No. 23A520056).

References

1. Zhai T, Liu JB, Wang DQ. 2023. Optimization path of agricultural products marketing channel based on innovative industrial chain. Econ Change Restruct. 56(6):3949-3977.
2. Xiong YL. 2023. Marketing strategy of characteristic agricultural products under the background of 'internet plus'. Prod Plann Control. 2023(3):1-12.
3. Liu QY, Zhao X, Shi KH. 2023. The analysis of agricultural internet of things product marketing by deep learning. J Supercomput. 79(4):4602-4621.
4. Nie WJ, Li TP, Zhu LQ. 2020. Market demand and government regulation for quality grading system of agricultural products in China. J Retail Consum Serv. 56:102134.
5. Xiao J, Wang WL, Tsai SB. 2021. Coupling of agricultural product marketing and agricultural economic development based on big data analysis and 'internet plus'. Mobile Inf Syst. 2021(8):1-10.
6. Wang HB, Gao J, Kang BH, Lyu P, Shi YX. 2022. Analysis and research on the marketing strategy of agricultural products based on artificial intelligence. Math Probl Eng. 2022:1-7.
7. Yang H, Zheng ZH, Sun CE. 2022. E-Commerce marketing optimization of agricultural products based on deep learning and data mining. Comput Intell Neurosci. 2022:6564014.
8. Huang C, Chen YL. 2021. Agricultural business and product marketing effected by using big data analysis in smart agriculture. Acta Agric Scand B Soil Plant Sci. 71(3):1-12.
9. Shi Y, Li XH, ZhouGong S, Li X, Wang H. 2022. Precise marketing classification of agricultural products for e-commerce live broadcast platform using clustering. Mobile Inf Syst. 2022(03):1-8.
10. Xu SW, Li DH, Zhang YE, Chen W, Zhuang JY, Liu JJ, *et al*. 2020. Development of portable user-interactive holographic information collector for agricultural product markets. Int J Agric Biol Eng. 13(3):143-153.
11. Tian T, Zhang YC, Mei YM. 2022. Intelligent analysis of precision marketing of green agricultural products based on big data and GIS. Earth Sci Inform. 15(3):1395-1406.
12. Yang YK. 2021. The application of big data in the analysis of factors affecting agricultural business and product marketing in the intelligent agricultural platform. Acta Agric Scand B Soil Plant Sci. 71(4):1-15.
13. Ma HY, Zhang XY. 2022. Construction of smart marketing model of agricultural products e-commerce in the era of big data. Mobile Inf Syst. 2022:3016554.
14. Chao S. 2022. Construction model of e-commerce agricultural product online marketing system based on blockchain and improved genetic algorithm. Secur Commun Netw. 107:4055698.
15. Gródek-Szostak Z, Malik G, Kajrunajtys D, Szeląg-Sikora A, Sikora J, Kuboń M, *et al*. 2019. Modeling the dependency between extreme prices of selected agricultural products on the derivatives market using the linkage function. Sustainability. 11(15):4144.
16. Dong LY. 2022. Analysis on influencing factors of consumer trust in e-commerce marketing of green agricultural products based on big data analysis. Math Probl Eng. 2022:8221657.
17. Lin FT. 2021. Study on the optimization of market competition of potato planting industry to agricultural products economic management based on intelligent system management. J Intell Fuzzy Syst. 40(4):5935-5944.
18. Luan GZ, Zhao F, Jia YW. 2022. Market analysis of characteristic agricultural products from the perspective of multi-source data: a case study of wild edible mushrooms. Sustainability. 14(21):14381.
19. Zia A, Alzahrani M. 2022. Investigating the effects of e-marketing factors for agricultural products on the emergence of sustainable consumer behaviour. Sustainability. 14(20):13072.
20. Graubner M, Salhofer K, Tribl C. 2021. A line in space: pricing, location, and market power in agricultural product markets. Annu Rev Resour Econ. 13:85-107.